



Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação

Projeto de Pesquisa – Iniciação Científica

Predição de Séries Temporais Baseada em Redes Neurais Artificiais

Submetido ao

Programa Institucional da Bolsas de Iniciação Científica (PIBIC)

Departamento de Engenharia de Computação e Automação Industrial (DCA)

Faculdade de Engenharia Elétrica e de Computação (FEEC)

Universidade Estadual de Campinas (UNICAMP)

CEP 13083-852, Campinas, São Paulo (SP)

Candidato: João Pedro De Oliveira Pagnan

RA: 199727

Orientador: Prof. Levy Boccato

Data de início: 08/2020

Duração: 12 meses

Campinas, 02 de Maio de 2020

Resumo

Neste trabalho, consideraremos o problema de predição de séries temporais baseada em redes neurais artificiais, que são estruturas bio-inspiradas de processamento dotadas de uma enorme capacidade de elaborar mapeamentos não-lineares. Serão consideradas a clássica rede MLP e algumas propostas de redes recorrentes, como as LSTM e as redes com estados de eco. Diferentes estratégias de aprendizado serão testadas em cenários de predição com características distintas, sendo o objetivo principal desse estudo o estabelecimento de bases para uma análise comparativa das metodologias, análise esta que tem um valor acadêmico e formativo significativo.

1 Introdução e Justificativa

A tarefa de prever valores futuros de uma série temporal a partir de valores presentes e passados é uma das aplicações mais instigantes em tratamento da informação [Principe et al., 1999]. Tal problema se reveste de uma importância decisiva numa gama de aplicações práticas que envolvem áreas tão distintas quanto econometria e automação industrial, ou meteorologia e biologia populacional. De certa forma, tal interesse prático se liga a dois anseios fundamentais: (1) o desejo de antecipar padrões de comportamento com a finalidade de estabelecer metodologias racionais de decisão, e (2) a busca por construir modelos adequados de fenômenos diversos, de modo a possibilitar uma maior compreensão acerca do comportamento dinâmico subjacente aos próprios dados, o que reveste a tarefa de uma importância teórica e prática ainda maior.

Tipicamente, o problema de predição é abordado por meio de um receituário com três etapas fundamentais (que não precisam necessariamente seguir uma ordem rígida) [Romano et al., 2011]: escolha de uma estrutura de filtragem que fará o papel de modelo de predição ou *preditor*, adoção de um critério que expresse a qualidade da predição efetuada pelo dispositivo escolhido e emprego de um método de otimização que, tendo em vista o binômio estrutura/critério adotado, forneça um conjunto ótimo de parâmetros livres a serem usados pelo preditor. A segunda dessas etapas é, em geral¹, definida a partir do uso de um critério de erro quadrático médio (EQM) de predição, que origina a seguinte função custo [Haykin, 2013]:

$$J_{\text{EQM}} = E\{(x(n+P) - y(n))^2\},$$

¹ Devem ser mencionados, não obstante, esforços bastante meritorios e promissores no sentido de empregar grandezas derivadas da teoria de informação no processo de elaboração de critérios de filtragem [Erdogmus e Principe, 2006].

sendo $E\{\cdot\}$ o operador *esperança matemática*, $y(n)$ a saída do preditor no instante n e $x(n + P)$ o valor da série analisada no instante $n + P$. Nesse caso, assume-se que a predição se dá P passos à frente e que o mapeamento expresso por $y(n)$ depende de $x(n)$ e, possivelmente, de valores passados.

Por outro lado, a definição da estrutura e do algoritmo de otimização adotados depende de uma reflexão maior acerca da natureza do problema em mãos. Embora seja comum (e, até certo ponto, bastante útil), inicialmente recorrer ao clássico ferramental linear, como é o caso do modelo autorregressivo (AR) [Haykin, 2013], é preciso, em muitos casos práticos, adotar estruturas de filtragem de cunho *fortemente não-linear*.

Nesse caso, surgem como candidatas interessantes ao papel de estrutura de predição as *redes neurais artificiais* [Haykin, 2008], dispositivos bio-inspirados de processamento cujo *modus operandi* se baseia em determinadas características do sistema nervoso. Consideramos particularmente relevantes as redes neurais do tipo MLP (do inglês *multi-layer perceptron*) [Bishop, 1996]. Em casos em que há a marcada necessidade de introdução de um caráter dinâmico, também devem ser levadas em conta as *redes recorrentes* [Connor et al., 1994] (RNNs, do inglês *recurrent neural networks*), que podem corresponder a versões das redes supracitadas em que há a presença de laços de realimentação.

Neste contexto, um modelo recorrente de destacada importância na literatura é a chamada *long short-term memory* (LSTM) [Greff et al., 2017], a qual apresenta um potencial de recordar valores dentro de sua estrutura relacionados a horizontes de tempo arbitrários. Em outras palavras, ela é capaz de memorizar informações tanto de curto quanto de longo prazo do histórico de eventos relacionados ao problema. Por sua vez, a GRU (*Gated Recurrent Unit*) também explora a noção de porta (*gate*) na definição do comportamento da unidade de processamento, mas de uma forma simplificada em relação à LSTM, obtendo também resultados promissores em algumas aplicações [Cho et al., 2014].

Outra perspectiva promissora com RNNs refere-se às redes neurais com estados de eco (ESNs, do inglês *echo state networks*). Estas estruturas combinam a capacidade de processamento dinâmico de RNNs com a simplicidade de treinamento típica de modelos lineares [Boccato, 2013].

Tendo em vista a diversidade de opções trazidas pela ideia de usar um *framework* baseado em redes neurais para a predição de séries, parece-nos bastante interessante conduzir um estudo em que as duas opções expostas sejam comparadas no âmbito de um conjunto cenários relevantes: esse é o objetivo central deste trabalho. O estudo, além de ter um significativo valor acadêmico, também propiciará ao candidato a oportunidade de adquirir uma formação sólida na área de redes neurais, bem como na linha da pesquisa de séries temporais como um todo.

2 Objetivos

O primeiro objetivo deste trabalho é a realização de um estudo teórico envolvendo sinais e sistemas, elementos de processos estocásticos, filtragem adaptativa supervisionada, redes neurais, aprendizado de máquina e elementos de modelamento / predição de séries temporais. Em particular, pretendemos cobrir algumas arquiteturas de redes neurais, tais como as redes *perceptron* de múltiplas camadas (MLP, do inglês *multilayer perceptron*), e os modelos recorrentes das LSTMs, GRUs e ESNs. Esse estudo, além de ter um papel fundamental no esquema da investigação associada ao projeto, será de grande importância para a formação de um jovem pesquisador em temas como processamento adaptativo de sinais e inteligência computacional.

Um segundo objetivo deste trabalho é a implementação e a análise sistemática das estruturas neurais no contexto da predição de diferentes tipos de séries - séries com características sazonais pronunciadas, séries caóticas e séries estocásticas (sinais de informação) – e, em cada caso, serão avaliados aspectos metodológicos desejáveis.

Finalmente, um terceiro objetivo do projeto é a redação de um relatório final que apresente as diferentes estruturas, algoritmos de ajuste de pesos e metodologias de seleção de variáveis de maneira didática e que contenha uma discussão ampla e representativa dos diversos ensaios realizados e dos resultados deles decorrentes. Desejamos produzir um documento que sirva de referência a futuros interessados na área, particularmente no seio do grupo no qual o aluno está inserido.

3 Metodologia

Na primeira etapa do projeto, será realizada uma ampla revisão bibliográfica formada por duas etapas essenciais, a primeira voltada ao estudo de sinais e sistemas [Oppenheim et al., 1996], processos estocásticos e filtragem adaptativa supervisionada [Haykin, 2013], e a outra dedicada às redes neurais selecionadas [Haykin, 2008, Géron, 2019]. Todas as etapas desse estudo serão acompanhadas de implementações práticas pertinentes.

Na segunda etapa, efetuar-se-á um planejamento cuidadoso dos diversos cenários de teste, dos processos de configuração das estruturas de processamento, dos algoritmos de aprendizado adotados e das metodologias de seleção de variáveis [Guyon e Elisseeff, 2003]. Também serão avaliadas as métricas de análise a serem adotadas de modo a permitir uma adequada comparação dos diversos *setups*. Após esse planejamento, o aluno estará apto a conduzir os testes e a compilar os resultados obtidos, os quais, a seguir, serão alvo de uma criteriosa análise. Ao longo da etapa, haverá sempre uma efetiva preocupação

com a validade e a representatividade das análises realizadas, para que a empreitada como um todo conduza a conclusões de valor prático.

Por fim, a terceira e final etapa será devotada ao processo de redação do relatório final, o qual deve apresentar de forma didática as diversas ferramentas empregadas e de maneira sistemática os resultados obtidos e análises empreendidas.

4 Material

Serão utilizados os recursos computacionais do DSPCom (Laboratório de Processamento de Sinais para Comunicações) e do Departamento de Engenharia de Computação e Automação Industrial da Faculdade de Engenharia Elétrica e de Computação da Unicamp. Estarão também disponíveis os recursos associados ao sistema de bibliotecas dessa universidade.

5 Plano de Trabalho / Metas Semestrais

As três etapas descritas na seção 3 serão desenvolvidas segundo o cronograma a seguir.

Etapa	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
Etapa 1	x	x	x	x	x	x						
Etapa 2					x	x	x	x	x	x		
Etapa 3									x	x	x	x

6 Síntese da Bibliografia Fundamental

[Attux, 2005] R. R. F. Attux, *Novos Paradigmas para Equalização e Identificação de Canais Baseados em Estruturas Não-Lineares e Algoritmos Evolutivos*, Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação – Unicamp, 2005.

[Bocato, 2013] L. Bocato, *Novas Propostas e Aplicações de Redes Neurais com Estados de Eco*, Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação – Unicamp, 2013.

- [Bishop, 2006] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [Cho et al., 2014] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, “*Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*”, arXiv:1406.1078, 2014.
- [Connor et al., 1994] J. T. Connor, R. Douglas Martin, L. E. Atlas, “*Recurrent Neural Networks and Robust Time Series Prediction*”, IEEE Transactions on Neural Networks, Vol. 5, No. 2, pp. 240-254, 1994.
- [Erdogmus e Principe, 2006] D. Erdogmus, J. Principe, “*From Linear Adaptive Filtering to Nonlinear Information Processing*”, IEEE Signal Processing Magazine, Vol. 23, No. 6, pp. 14-33, 2006.
- [Géron, 2019] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*, O'Reilly Media, 2^a ed., 2019.
- [Greff et al., 2017] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, “*LSTM: A Search Space Odyssey*”, IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222-2232, 2017.
- [Guyon e Elisseeff, 2003] I. Guyon, A. Elisseeff, “*An Introduction to Variable and Feature Selection*”, Journal of Machine Learning Research, No. 3, pp. 1157-1182, 2003.
- [Haykin, 2008] S. Haykin, *Neural networks and Learning Machines*, Prentice Hall, 3^a ed., 2008.
- [Haykin, 2013] S. Haykin, *Adaptive Filter Theory*, Pearson, 5^a ed., 2013.
- [Oppenheim et al., 1996] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, *Signals and Systems*, Prentice Hall, 1996.

[Principe et al., 1999] J. C. Principe, N. R. Euliano, W. C. Lefebvre, *Neural and Adaptive Systems*, Wiley, 1999.

[Romano et al., 2011] J. M. T. Romano, R. Attux, C. C. Cavalcante e R. Suyama. *Unsupervised signal processing: channel equalization and source separation*. CRC Press, 2011.